

Chapter 13

Heavy-tailed random matrices

Z. BURDA and J. JURKIEWICZ

Marian Smoluchowski Institute of Physics and
Mark Kac Complex Systems Research Center,
Jagiellonian University,
Reymonta 4, 30-059 Kraków, Poland

Abstract

We discuss non-Gaussian random matrices whose elements are random variables with heavy-tailed probability distributions. In probability theory heavy tails of the distributions describe rare but violent events which usually have dominant influence on the statistics. They also completely change universal properties of eigenvalues and eigenvectors of random matrices. We concentrate here on the universal macroscopic properties of (1) Wigner matrices belonging to the Lévy basin of attraction, (2) matrices representing stable free random variables and (3) a class of heavy-tailed matrices obtained by parametric deformations of standard ensembles.

13.1 Introduction

Gaussian random matrices have been studied over many decades and are well known by now [Meh04]. Much less is known about matrices whose elements display strong fluctuations described by probability distributions with heavy tails.

Probably the simplest example of a matrix from this class is a real symmetric ($A_{ij} = A_{ji}$) random matrix A_N with elements A_{ij} , $1 \leq i \leq j \leq N$, being

independent identically distributed (i.i.d.) centered real random variables with a probability density function (p.d.f.) falling off as a power

$$p(x) \sim |x|^{-\alpha-1} . \quad (13.1.1)$$

for $|x| \rightarrow \infty$. The smaller is the value α the heavier is the tail. Higher moments of this distribution do not exist. For $\alpha \in (0, 1]$ the tail is extremely heavy and the mean-value does not exist since the corresponding integral $\int xp(x)dx$ is divergent. For $\alpha \in (1, 2]$ the mean-value does exist but the variance does not. The influence of heavy tails on the statistical properties of a random matrix is enormous. It is particularly apparent for $\alpha < 1$. In this case the elements A_{ij} assume values scattered over a wide range which itself quickly increases when N goes to infinity. The largest element $|a_{max}|$ of the matrix is of the order $|a_{max}| \sim N^{2/\alpha}$ and its value strongly fluctuates from matrix to matrix. The distribution of the normalized value of the maximal element $x = |a_{max}|/N^{2/\alpha}$ is given by a Fréchet distribution which itself has a heavy-tail. The largest element of the matrix may be larger than the sum of all remaining ones. The values of the elements change so dramatically from matrix to matrix that one cannot speak about a typical matrix or about self-averaging for large N . In the limit $N \rightarrow \infty$ matrices A_N may look effectively very sparse¹. Indeed if one considers a rescaled matrix $A_N/|a_{max}|$ one will find that only a finite fraction of all elements of this matrix will be significantly different from zero. This effective sparseness is quantified by the inverse participation ratio Y_2 constructed from normalized weights $w_{ij} = |a_{ij}|/\sum_{ij} |a_{ij}|$, which sum up to unity $\sum_{ij} w_{ij} = 1$,

$$Y_2 = \frac{2}{N(N+1)} \overline{\sum_{i \leq j} w_{ij}^2} . \quad (13.1.2)$$

The bar denotes the average over matrices A_N . In the limit $N \rightarrow \infty$ the participation ratio is $Y_2 = 1 - \alpha > 0$ for $\alpha \in (0, 1)$ [Bou97]. This means that only a finite fraction of matrix elements is relevant in a given realization of the matrix. This is a completely different behavior than the one known from considerations of Gaussian random matrices. For $\alpha \leq [1, 2)$, although $Y_2 = 0$ in the limit $N \rightarrow \infty$, one still observes very large fluctuations of individual matrix elements which in particular lead to a localization of eigenvectors that will be shortly discussed towards the end of the next section. Only for $\alpha > 2$ the behavior of matrices resembles² that known for Gaussian matrices. In this case, the variance σ^2 of (13.1.1) is finite and the eigenvalue density of the matrix A_N/\sqrt{N} converges for $N \rightarrow \infty$ to the Wigner semicircle law $\rho(\lambda) =$

¹ Sparse random matrices are discussed in Chapter 23 of this book.

²The convergence to the limiting semicircle law is generically very slow in the presence of power-law tails and moreover for $\alpha \leq 4$ microscopic properties are significantly different than for generic Gaussian matrices as we will shortly mention later.

$\sqrt{4\sigma^2 - \lambda^2}/(2\pi\sigma^2)$, independently of the details of the probability distribution. Random matrices having the same limiting eigenvalue density for $N \rightarrow \infty$ are said to belong to the same macroscopic universality class. For $\alpha > 2$ it is called a Gaussian universality. This class is very broad and comprises a whole variety of random matrices. In particular one can prove [Pas72] that if A_N is a symmetric random matrix with independent (but not necessarily identically distributed) centered entries with the same variance σ^2 , then the condition for the eigenvalue distribution of A_N/\sqrt{N} to converge to the Wigner semicircle law is

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i \leq j} \int_{|x| > \epsilon\sqrt{N}} x^2 p_{ij}(x) dx = 0 \quad (13.1.3)$$

where ϵ is any positive number and $p_{ij}(x)$ is the p.d.f. for the ij -th element of the matrix. As a matter of fact this condition is almost identical as the Lindeberg condition known from the central limit theorem for the distribution of a sum of random numbers to converge to a normal distribution [Fel71]. The fastest convergence to the limiting semicircle law is achieved for matrices whose elements are independent Gaussian random variables. A prominent place in this macroscopic universality class is taken by the ensemble of symmetric Gaussian matrices whose diagonal elements have a twice bigger variance than the off-diagonal ones $\mathcal{N}(0, \sigma^2(1 + \delta_{ij}))$. Clearly, such matrices fulfill the Lindeberg condition. The probability measure in the ensemble of such matrices can be written as

$$d\mu(A) = DA \exp -\frac{1}{2\sigma^2} \text{tr} A^2 \quad (13.1.4)$$

where DA is a flat measure $DA = \prod_{1 \leq i \leq j \leq N} dA_{ij}$. The measure $d\mu(A)$ is manifestly invariant with respect to the orthogonal transformations: $A \rightarrow OAO^T$, where O is an orthogonal matrix. This is the GOE ensemble. In the limit $N \rightarrow \infty$ the eigenvalue density of A_N/\sqrt{N} approaches a well-known semicircle distribution. In a similar way one can construct GUE and GSE ensembles which are extensively discussed in other chapters of the book. In this chapter we will be mostly interested in matrices for which the variance does not exist.

13.2 Wigner-Lévy matrices

In this section we will discuss properties of heavy tailed symmetric matrices with i.i.d. elements (13.1.1) for $\alpha \in (0, 2)$. We call them Wigner-Lévy matrices since the Lévy distribution is the corresponding stable law for $0 < \alpha < 2$ which plays an analogous role from the point of view of the central limit theorem as the Gaussian distribution for $\alpha \geq 2$ [Gne68]. The Lévy distributions are sometimes called α -stable laws.

Before we discuss Wigner-Lévy matrices let us briefly recall basic facts about Lévy distributions. In addition to a stability index α these laws are character-

ized by an asymmetry parameter $\beta \in [-1, 1]$, which will be discussed below, and a scale parameter $R > 0$, called the range, which plays a similar role as the standard deviation σ for the normal law. The statement that a Lévy distribution is a stable law (with respect to addition) means that a sum of two independent Lévy random variables $x = x_1 + x_2$ with a given index α is again a Lévy random variable with the same α . The range R and asymmetry β of the resulting distribution can be calculated as

$$R^\alpha = R_1^\alpha + R_2^\alpha, \quad \beta = \frac{\beta_1 R_1^\alpha + \beta_2 R_2^\alpha}{R_1^\alpha + R_2^\alpha}. \quad (13.2.1)$$

For $\beta_1 = \beta_2$ the asymmetry is preserved and the relation for the effective range is a generalization of the corresponding one for independent Gaussian random variables, where the sum is also a Gaussian random variable with the variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Actually for $\alpha = 2$ the range and the standard deviation are related as $\sigma = \sqrt{2}R$.

The p.d.f. $L_\alpha^{R,\beta}(x)$ of the Lévy distribution with the stability index α , the asymmetry β and the range R is conventionally written as a Fourier transform of the characteristic function, since its form is known explicitly. There are several definitions used in the literature, here we quote one, which seems to be the most common [Gne68, Nol10]

$$L_\alpha^{R,\beta}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk \widehat{L}_\alpha^{R,\beta}(k) e^{-ikx} \quad (13.2.2)$$

where³

$$c(k) = \ln \widehat{L}_\alpha^{R,\beta}(k) = -R^\alpha |k|^\alpha (1 + i\beta \operatorname{sgn}(k) \tan(\pi\alpha/2)) . \quad (13.2.3)$$

The logarithm of the characteristic function $c(k)$ is usually called a cumulant-generating function. This name is slightly misleading since for Lévy distributions only the first cumulant exists for $\alpha \in (1, 2)$ or none for $\alpha \in (0, 1]$. Therefore we will rather call it the c -transform. The characteristic function is known explicitly in contrast to the corresponding p.d.f. $L_\alpha^{R,\beta}(x)$ which can be expressed in terms of simple functions only for $\alpha = 1$, $\beta = 0$ (Cauchy distribution) $\alpha = 3/2$, $\beta = \pm 1$ (Smirnov distribution). For $\alpha = 2$, the characteristic function (13.2.3) becomes a Gaussian function independent of β . A stability of the Lévy distribution can be easily verified. A p.d.f. for the sum of two independent variables $x_{1+2} = x_1 + x_2$ is a convolution of the p.d.f.'s for individual components x_1 and x_2 and thus the corresponding characteristic function is a product of two characteristic functions. It is easy to check by inspection of (13.2.3) that the relations (13.2.1) are indeed satisfied. It is less trivial to demonstrate that the Fourier transform of the characteristic function (13.2.2)

³For $\alpha = 1$ it assumes a slightly form: $c(k) = -R|k| (1 + i(2\beta/\pi)\operatorname{sgn}(k) \ln(Rk))$.

is a non-negative function. Actually this is only the case for $0 < \alpha \leq 2$. Lévy distributions $L_\alpha^{R,\beta}(x)$ for $\alpha \in (0, 2]$ are the only stable laws. The asymmetry parameter β controls the skewness of the distribution. For $\beta = 0$, the characteristic function (13.2.3) is even and so is the p.d.f. $L_\alpha^{R,0}(x) = L_\alpha^{R,0}(-x)$. For $\beta \neq 0$ the p.d.f. is skew and has a different left and right asymptotic behavior

$$L_\alpha^{R,\beta}(x) \xrightarrow{x \rightarrow \pm\infty} (1 \pm \beta) \frac{\gamma_\alpha R^\alpha}{|x|^{\alpha+1}} \quad (13.2.4)$$

where $\gamma_\alpha = \Gamma(1 + \alpha) \sin(\pi\alpha/2)/\pi$. In the extreme cases $\beta = \pm 1$ one of the tails is suppressed and for $\alpha < 1$ the distribution becomes fully asymmetric with a support only on the positive (resp. negative) semiaxis. One should note that the mean value of the Lévy distribution for $\alpha > 1$ equals zero, independently of the asymmetry β . For $\alpha = 2$ the dependence on β disappears. If one sets $R = 1$ in (13.2.3) one obtains a standardized Lévy distribution. A Lévy random variable x with an arbitrary R can be obtained from the corresponding standardized one x_* by a rescaling $x = Rx_*$, hence $L_\alpha^{R,\beta}(x) = L_\alpha^{1,\beta}(x/R)/R$.

Let us now consider a symmetric matrix A_N with elements A_{ij} for $1 \leq i \leq j \leq N$ being i.i.d. Lévy random variables with the p.d.f. $L_\alpha^{R,\beta}(x)$. We are now interested in the eigenvalue density of such a matrix in the limit $N \rightarrow \infty$. As we shall see below the eigenvalue density of the matrix $A_N/N^{1/\alpha}$ converges to a limiting density $\rho(\lambda)$ which is completely different from the Wigner semicircle law and has an infinite support and heavy tails. The choice of the scaling factor $N^{1/\alpha}$ is related to the universal scaling properties of the Lévy distribution and for this reason can be viewed as a generalization of the scaling for Gaussian random matrices ($\alpha = 2$). The problem of the determination of the limiting eigenvalue distribution of Wigner-Lévy matrices is not simple because the standard methods used for Gaussian matrices or matrices from invariant ensembles do not apply here. A special method tailored to the specific universal features of heavy-tailed distributions was necessary to attack this problem. Such a method, being a beautiful adaptation of the cavity method which maximally exploits universal properties of α -stable distributions, was invented in [Ciz94]. The description of the method is beyond the scope of this chapter and we refer the interested reader to the original paper [Ciz94] and to the paper [Bur07] where the derivation was explained step by step and some details were corrected. Here we only quote the final result. The eigenvalue density $\rho(\lambda)$ of a Wigner-Lévy matrix is given by a Lévy function with an index $\alpha/2$, being a half of the index α of the p.d.f. $L_\alpha^{R,\beta}(x)$ used to generate the matrix elements, and an effective “running” range $\widehat{R}(\lambda)$ and asymmetry parameter $\widehat{\beta}(\lambda)$

$$\rho(\lambda) = \frac{1}{\Lambda} \widehat{\rho}\left(\frac{\lambda}{\Lambda}\right) \quad \text{where} \quad \widehat{\rho}(\lambda) = L_{\alpha/2}^{\widehat{R}(\lambda), \widehat{\beta}(\lambda)}(\lambda) \quad (13.2.5)$$

and

$$\Lambda = R \left(\frac{\Gamma(1 + \alpha) \cos(\pi\alpha/4)}{\Gamma(1 + \alpha/2)} \right)^{1/\alpha}. \quad (13.2.6)$$

The scale parameter Λ is proportional to the original range R . The functions $\hat{R}(\lambda)$, $\hat{\beta}(\lambda)$ satisfy a set of integral equations

$$\hat{R}^{\frac{\alpha}{2}}(\lambda) = \int_{-\infty}^{+\infty} dx |x|^{-\frac{\alpha}{2}} L_{\alpha/2}^{\hat{R}(x), \hat{\beta}(x)}(\lambda - x) \quad (13.2.7)$$

$$\hat{\beta}(\lambda) = \frac{\int_{-\infty}^{+\infty} dx \operatorname{sign}(x) |x|^{-\frac{\alpha}{2}} L_{\alpha/2}^{\hat{R}(x), \hat{\beta}(x)}(\lambda - x)}{\int_{-\infty}^{+\infty} dx |x|^{-\frac{\alpha}{2}} L_{\alpha/2}^{\hat{R}(x), \hat{\beta}(x)}(\lambda - x)} \quad (13.2.8)$$

where the integrals should be interpreted as Cauchy principal values. The equation (13.2.8), which was derived in [Bur07], is a corrected version of this equation given in [Ciz94].

We conclude the section with some comments.

The dependence on λ on the right hand side of (13.2.5) appears not only through the main argument of the function but also through the dependence of the effective parameters $\hat{R}(\lambda)$ and $\hat{\beta}(\lambda)$ on λ . This makes the resulting expression very complex. The equations for $\hat{\beta}(\lambda)$ and $\hat{R}(\lambda)$ cannot be solved analytically. It is also not easy to solve them numerically because the computation of the function $L_{\alpha}^{R, \beta}(x)$, which is the main building block of the above integral equations, is numerically unstable if one does the Fourier integral (13.2.2) in a straightforward way since its integrand is a strongly oscillating function. It is necessary to apply a non-trivial transformation of the integration contour in the complex plane to assure that the integration becomes numerically stable [Nol10]. We skip the details here and again refer the interested reader to [Bur07] for details.

The eigenvalue density (13.2.5) holds for any $\alpha \in (1, 2)$ and is independent of the asymmetry β of the p.d.f. for matrix elements. The independence of β might be surprising but it can be easily checked by inspection. Indeed, equations (13.2.7) and (13.2.8) depend only on α and do not involve any dependence on the parameters β or R of the p.d.f. for matrix elements, $L_{\alpha}^{R, \beta}(x)$. Numerically the solution (13.2.5) seems to extend also to the range of $\alpha \in (0, 1]$ but for symmetric distributions ($\beta = 0$) only.

The limiting eigenvalue density $\rho(\lambda)$ is an unimodal even function. The height of the maximum at $\lambda = 0$ is

$$\rho(0) = \frac{\Gamma(1 + 2/\alpha)}{\pi R} \left(\frac{\Gamma^2(1 + \alpha/2)}{\Gamma(1 + \alpha)} \right)^{1/\alpha}. \quad (13.2.9)$$

For large $|\lambda| \rightarrow \infty$ the eigenvalue density has heavy tails

$$\rho(\lambda) \sim \frac{1}{\pi} \Gamma(1 + \alpha) \sin\left(\frac{\alpha\pi}{2}\right) \frac{R^{\alpha}}{|\lambda|^{\alpha+1}} \quad (13.2.10)$$

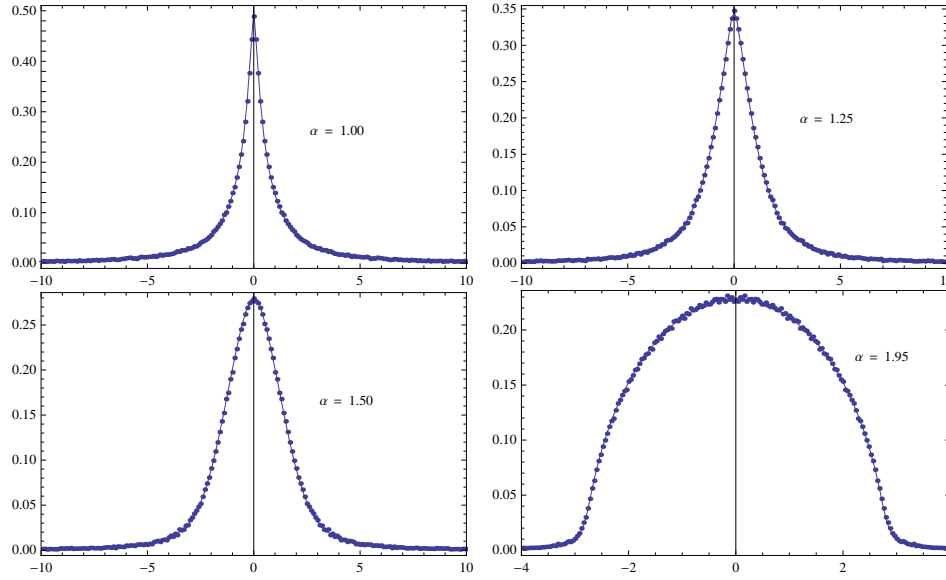


Figure 13.1: The eigenvalue density for infinite Wigner-Lévy matrices with the index $\alpha = 1.0, 1.25, 1.5, 1.95$ and the range $R = 1$ (solid line) and the corresponding numerical densities obtained by Monte-Carlo generated matrices of size 200×200 with entries distributed according to $L_\alpha^{1,0}(x)$.

with the same power as the p.d.f. for the matrix elements, although at first sight one might have the impression that the power should rather be $\alpha/2$ for it is the index of the Lévy function on the right hand side of (13.2.5). This is not the case because also the dependence of the running parameters $\hat{R}(\lambda)$ and $\hat{\beta}(\lambda)$ on λ contributes to a net asymptotic behavior. In Figure 13.1 we show limiting distributions $\rho(\lambda)$ of the Wigner-Lévy matrices with $R = 1$ and different values of α and compare them to eigenvalue histograms obtained numerically by Monte-Carlo generation of symmetric 200×200 matrices with i.i.d. entries with the p.d.f. $L_\alpha^{1,0}(x)$. The agreement is very good and finite size effects are small.

Another feature of Wigner-Lévy matrices which distinguishes them from Wigner matrices from the Gaussian universality class is a localization of the eigenvectors [Ciz94]. The degree of localization is measured by the inverse participation ratio

$$y_2 = \sum_{i=1}^N \psi_i^4 \quad (13.2.11)$$

calculated for the elements ψ_i 's of a normalized eigenvector $\sum_{i=1}^N \psi_i^2 = 1$. For Wigner matrices from the Gaussian universality class $y_2 = 0$ in the limit $N \rightarrow \infty$ and it is independent of the corresponding eigenvalue. For Wigner-Lévy

matrices with the index $\alpha \in (1, 2)$, the average inverse participation ratio $y_2 = y_2(\lambda)$ depends on the eigenvalue λ . For λ smaller than a certain critical value, $\lambda \leq \lambda_{cr}$, $y_2(\lambda) = 0$, however for $\lambda > \lambda_{cr}$ above this value $y_2(\lambda)$ is positive and it grows monotonically to one when λ increases [Ciz94]. This means that only a finite number of elements ψ_i are significantly different from zero or phrasing it differently that the vector is localized on a subset of cardinality of the order $1/y_2(\lambda)$. When λ goes to infinity $y_2(\lambda)$ tends to one. This means that for extremely large eigenvalues all but one elements of the corresponding vector are equal zero and the eigenvector is localized on exactly one state. The critical value λ_{cr} depends on α and grows monotonically from zero for $\alpha = 1$ to infinity for $\alpha = 2$. The fact that $\lambda_{cr} = 0$ for $\alpha = 1$ means that in this case (and also for $\alpha < 1$) all eigenvectors are localized.

The eigenvalue density (13.2.5) defines a whole universality class of Wigner-Lévy matrices. It is a counterpart of the Wigner semicircle law. The eigenvalue density of a matrix $A_N/N^{1/\alpha}$ with i.i.d. entries with a p.d.f. $p(x)$ will converge to the same limiting law if the function $p(x)$ has the same asymptotic behavior as $L_\alpha^{R,\beta}(x)$. More precisely, if the p.d.f. $p(x)$ is centered ($\int_{-\infty}^{+\infty} dx xp(x) = 0$) for $\alpha \in (1, 2)$ (or even for $\alpha \in (0, 1]$) and has the following asymptotic behavior

$$p(x) \xrightarrow{x \rightarrow \pm\infty} \frac{C_\pm}{|x|^{1+\alpha}} \quad (13.2.12)$$

then the eigenvalue distribution of the matrix $A_N/N^{1/\alpha}$ converges to the same limiting density $\rho(x)$ (13.2.5) as for the corresponding matrix with the p.d.f. $L_\alpha^{R,\beta}(x)$ with the same index α and

$$R = \left(\frac{C_+ + C_-}{2\gamma_\alpha} \right)^{1/\alpha}, \quad \beta = \frac{C_+ - C_-}{C_+ + C_-}, \quad (13.2.13)$$

where as before $\gamma_\alpha = \Gamma(1 + \alpha) \sin(\pi\alpha/2)/\pi$. One can probably extend this macroscopic universality to a class of matrices with independent but not necessarily identically distributed entries however having distributions with the same asymptotic behavior.

We finish this section with a comment on microscopic properties of Wigner matrices with heavy tails (13.1.1). We have seen so far that macroscopic properties of Wigner matrices change at $\alpha = 2$. For $\alpha > 2$ Wigner matrices belong to the Gaussian universality class and their eigenvalue density converges for large N to the Wigner semicircle law while for $\alpha < 2$ they belong to the Lévy universality class and their eigenvalue density converges to the limiting distribution given by (13.2.5). One may ask if $\alpha = 2$ is also a critical value for microscopic properties like for instance eigenvalue correlations or the statistics of the largest eigenvalue λ_{max} . This question has already been partially studied. It was found that in this case the critical value of the exponent α

is rather $\alpha = 4$ [Bir07a]. For $\alpha > 4$ the largest eigenvalue of the matrix A_N/\sqrt{N} , for $N \rightarrow \infty$, fluctuates around the upper edge of the support of the Wigner semicircle distribution and the fluctuations are of the order $N^{-2/3}$. A rescaled quantity $x = (\lambda_{\max} - \lambda_{\text{edge}})N^{2/3}$ obeys the Tracy-Widom statistics [Tra94, Tra96] although the convergence to this limiting distribution is rather slow. For $\alpha < 4$ the largest eigenvalue is of the order $N^{(4-\alpha)/(2\alpha)}$ and a rescaled quantity $y = \lambda_{\max}/N^{(4-\alpha)/(2\alpha)}$ is distributed according to a modified Fréchet law. For $\alpha = 4$ which is a marginal case the two regimes are mixed in the proportions depending on details of the p.d.f. for matrix elements, in particular on the amplitude of the tail. Roughly speaking for $\alpha > 4$ the eigenvalue repulsion shapes the microscopic properties of the matrix and leads to the Tracy-Widom statistics while for $\alpha < 4$ the repulsion plays a secondary role. The dominant effect is in this case related to fluctuations in the tail of the distribution which are so large that the repulsion can be neglected and the largest eigenvalues can be treated as independent of each other. Actually it has been known for some time [Sos04] that indeed the largest eigenvalues are given by a Poisson point process with a Fréchet intensity related to the statistics of the largest elements in the random matrix coming from the tail of the distribution (13.1.1) for $\alpha \leq 2$. In paper [Bir07a] an argument was given that basically the same picture holds also for $2 < \alpha < 4$.

Wigner matrices are discussed in Chapter 21. The interested reader can find there a discussion also of other aspects of this class of random matrices.

13.3 Free random variables and free Lévy matrices

It is sometimes convenient to think of whole matrices as entities and to formulate for them probabilistic laws. One can ask for example if one can calculate the eigenvalue density $\rho_{1+2}(\lambda)$ of a sum of two symmetric (or hermitian) $N \times N$ random matrices

$$A_{1+2} = A_1 + A_2 \tag{13.3.1}$$

given the densities $\rho_1(\lambda)$ and $\rho_2(\lambda)$ of A_1 and A_2 . In general the answer to this question is negative since the resulting distribution depends on many other factors. The situation becomes less hopeless for large matrices. It turns out that for $N \rightarrow \infty$, $\rho_{1+2}(\lambda)$ depends only on $\rho_1(\lambda)$ and $\rho_2(\lambda)$ if A_1 and A_2 are independent random matrices or saying more precisely if they are free. The freeness is a concept closely related to the independence. The independence itself is not sufficient. The freeness additionally requires a complete lack of angular correlations. Such correlations may appear even for independent matrices if they are generated from a matrix-ensemble which hides some characteristic angular pattern. An example is just the ensemble of Wigner-Lévy matrices whose probability measure is not rotationally invariant. This measure favors some specific

angular directions. In effect, even if one picks at random two matrices from this ensemble, they both will prefer some angular directions and thus will have some sort of mutual correlations. One can remove the correlations by a uniform angular randomization of the matrices A_1 and A_2

$$A_{1\boxplus 2} = O_1 A_1 O_1^T + O_2 A_2 O_2^T \quad (13.3.2)$$

where O_i 's are random orthogonal matrices with a uniform probability measure in the group of orthogonal matrices⁴. The matrix $O_i A_i O_i^T$ has exactly the same eigenvalue content as A_i . Actually to achieve the effect of the angular decorrelation it is sufficient to rotate only one of A_i 's. This type of addition is called a free addition and we denote it by \boxplus . Of course if A_i 's are generated from an ensemble with a rotationally invariant probability measure, as for instance $DA \exp -\text{tr} V(A)$, then the two types of additions (13.3.1) and (13.3.2) are identical and $\rho_{1+2}(\lambda) = \rho_{1\boxplus 2}(\lambda)$. Otherwise they are different and $\rho_{1+2}(\lambda) \neq \rho_{1\boxplus 2}(\lambda)$. For example a sum of independent identically distributed Wigner-Lévy matrices (13.3.1) is a Wigner-Lévy matrix which is not rotationally invariant while a free sum of Wigner-Lévy matrices is a rotationally invariant matrix, which has a different eigenvalue density, so in this case $\rho_{1+2}(\lambda) \neq \rho_{1\boxplus 2}(\lambda)$. Generally, for large matrices the eigenvalue density $\rho_{1\boxplus 2}(\lambda)$ of a free sum depends only on $\rho_1(\lambda)$ and $\rho_2(\lambda)$ and it can be uniquely determined from them [Voi92].

A theory of free addition was actually developed in probability theory of non-commutative objects (operators) long before random matrices entered the scene [Voi85]. It was originally formulated in terms of von Neumann algebras equipped with a trace-like normal state τ , which was introduced to generalize the concept of uncorrelated variables known from a classical probability. More specifically in classical probability two real random variables x_1 and x_2 are uncorrelated if the correlation function, calculated as the expectation value $E(\hat{x}_1 \hat{x}_2) = 0$ for centered variables, $\hat{x}_i = x_i - E(x_i)$, vanishes. In free probability, by analogy, elements X_i are free if $\tau(\hat{X}_{\pi(1)} \hat{X}_{\pi(2)} \dots \hat{X}_{\pi(m)}) = 0$ for any permutation π of the corresponding centered elements $\hat{X}_i = X_i - \tau(X_i) \cdot 1$. A link to large matrices was discovered later [Voi91] when it was realized that a non-commutative probability space with free elements X_i 's can be mapped onto a set of large $N \times N$ matrices, $N \rightarrow \infty$, of the form $X_i = U_i D_i U_i^\dagger$, where D_i are diagonal real matrices and U_i random unitary (or orthogonal, O_i) matrices distributed with a uniform probability measure on the group. In this mapping the state function τ corresponds to a standard normalized trace operation $\tau(\dots) \leftrightarrow \frac{1}{N} \text{tr}(\dots)$. This observation turned out to be very fruitful both for free probability and for theory of large random matrices since one could successfully apply results of a free probability to random matrices and vice versa

⁴For hermitian matrices A_i one uses unitary rotations $A_{1+2} = U_1 A_1 U_1^\dagger + U_2 A_2 U_2^\dagger$

[Voi92]. Free probability and its relation to random matrices and planar combinatorics for the case when all moments of the probability distribution exist [Spe94, Nic06] is discussed in detail in Chapter 22. In this section we concentrate on heavy-tailed free distributions for which higher moments do not exist. In particular we shall discuss free stable laws and their matrix realizations. Before we do that we recall basic concepts to make the discussion of this section self-contained.

Actually the law of addition of free random variables is in many respects analogous to the law of addition of independent real random variables. We will therefore begin by briefly recalling the law of addition in classical probability and by analogy describe the corresponding steps in free probability. The p.d.f. $p_{1+2}(x)$ of a sum $x = x_1 + x_2$ of independent real random variables x_1 and x_2 is given by a convolution of the individual p.d.f.'s $p_{1+2}(x) = (p_1 * p_2)(x)$. Thus, the characteristic function $\widehat{p}(k) = \int dk e^{ikx} p(x)$ for the sum is a product of the corresponding characteristic functions $\widehat{p}_{1+2}(k) = \widehat{p}_1(k) \widehat{p}_2(k)$. This is more conveniently expressed in terms of a cumulant-generating function (c -transform) defined as $c(k) = \ln \widehat{p}(k)$, as a simple additive law

$$c_{1+2}(k) = c_1(k) + c_2(k). \quad (13.3.3)$$

It turns out that one can find a corresponding object in free probability [Voi92], a free cumulant-generating function alternatively called R -transform, for which the free addition (13.3.2) leads to a corresponding additive rule

$$R_{1\boxplus 2}(z) = R_1(z) + R_2(z). \quad (13.3.4)$$

For a given eigenvalue density $\rho(\lambda)$ one defines a moment-generating function as a Cauchy transform of the eigenvalue density [Voi92]

$$G(z) = \int_{-\infty}^{+\infty} \frac{\rho(\lambda) d\lambda}{z - \lambda} \quad (13.3.5)$$

known also as the resolvent or Green function. The free-cumulant-generating function (R -transform) is related to the Green function as

$$z = G \left(R(z) + \frac{1}{z} \right). \quad (13.3.6)$$

The last equation can be inverted for $G(z)$

$$z = R(G(z)) + \frac{1}{G(z)}. \quad (13.3.7)$$

In short $z \rightarrow R(z) + 1/z$ is the inverse function of $z \rightarrow G(z)$ and thus for a given resolvent one can determine the R -transform and vice versa. Using the

addition law (13.3.4) one can now give a step-by-step algorithm to calculate the eigenvalue density $\rho_{1\boxplus 2}(\lambda)$ of the free sum (13.3.2) from $\rho_1(\lambda)$ and $\rho_2(\lambda)$. First one determines the resolvents $G_1(z)$ and $G_2(z)$ using (13.3.5), then the corresponding R -transforms $R_1(z)$ and $R_2(z)$ using (13.3.6) and finally $R_{1\boxplus 2}(z)$ using the addition law (13.3.4). Having found $R_{1\boxplus 2}(z)$ one proceeds in the opposite order. One reconstructs the corresponding resolvent $G_{1\boxplus 2}(z)$ (13.3.7) and then the density $\rho_{1\boxplus 2}(\lambda)$ by the inverse of (13.3.5)

$$\rho(\lambda) = -\frac{1}{\pi} \text{Im} G(\lambda + i0^+) \quad (13.3.8)$$

which follows from the relation $(x + 0^+)^{-1} = P.V.x^{-1} - i\pi\delta(x)$. This completes the task of calculating the eigenvalue density of a free sum $\rho_{1\boxplus 2}(\lambda)$ from $\rho_1(\lambda)$ and $\rho_2(\lambda)$. This procedure can be fully automatized and it actually has been implemented for a certain class of matrices [Rao06].

The correspondence between the laws of addition (13.3.3) and (13.3.4) and between the logical structures behind these laws in classical and free probability has also profound theoretical implications. One of them is a bijection between infinitely divisible laws of classical probability and the laws in free probability [Ber93]. Using this bijection one can derive the R -transform for stable laws in free probability⁵ [Ber99]

$$R(z) = \begin{cases} bz^{\alpha-1} & \text{for } \alpha \in (0, 2] \text{ and } \alpha \neq 1 \\ -i(1+\beta) - (2\beta/\pi) \ln z & \text{for } \alpha = 1 \end{cases} \quad (13.3.9)$$

where $b = -e^{i\alpha(1+\beta)\pi/2}$ for $\alpha \in (0, 1)$ and $b = e^{i(\alpha-2)(1+\beta)\pi/2}$ for $\alpha \in (1, 2]$. The stable R -transforms (13.3.9) are in one-to-one correspondence with the c -transforms of Lévy distributions (13.2.3) and they fully classify all free stable laws and allow one to determine the free probability densities for these stable laws. These densities are equal to the eigenvalue densities of free Lévy matrices being matrix realizations of the stable free random variables.

Let us illustrate how this procedure works for a stable law with the stability index $\alpha = 2$ and unit range. Using (13.3.9) we have $R(z) = z$. Inserting this to (13.3.7) we obtain an equation for the resolvent $G(z) = z - 1/G(z)$ which gives in the upper complex half-plane $G(z) = (z - \sqrt{z^2 - 4})/2$. Finally using (13.3.8) we find a Wigner law $\rho(\lambda) = \sqrt{4 - \lambda^2}/2\pi$ with $\sigma = 1$. We see that the Wigner law is equivalent in free probability to the normal law in classical probability. Secondly consider the case for $\alpha = 1$, $\beta = 0$. Proceeding in the same way as above we find respectively

$$R(z) = -i, \quad G(z) = \frac{1}{z+i}, \quad \rho(\lambda) = \frac{1}{\pi} \frac{1}{1+\lambda^2}. \quad (13.3.10)$$

⁵We give only the standardized version which corresponds to the unit range. An R -transform with a range r can be obtained from the standardized one by a rescaling $R_r(z) = rR(rz)$.

This case is special because the stable density is identical in classical and free probability. For other values of α one can find the free stable laws by applying the equation (13.3.7) to the R -transform (13.3.9) which gives the following equation for the resolvent

$$bG^\alpha(z) - zG(z) + 1 = 0. \quad (13.3.11)$$

This equation can be solved analytically for a couple of values of the parameter α for which it is just a quadratic, cubic or quartic equation. The solution can be then used to calculate the eigenvalue density (13.3.8). For other values the eigenvalue density can be determined numerically.

The equation (13.3.11) may be used to extract the asymptotic behavior of the corresponding eigenvalue density (13.3.8). We will give the result only for the symmetric case ($\beta = 0$) and for the range $R = 1$. For small eigenvalues $|\lambda| \rightarrow 0$ it reads [Bur02]

$$\rho(\lambda) = \frac{1}{\pi} \left(1 - \frac{3-\alpha}{2\alpha^2} \lambda^2 + \dots \right) \quad (13.3.12)$$

while for large ones, $|\lambda| \rightarrow \infty$

$$\rho(\lambda) \sim \frac{1}{\pi} \sin\left(\frac{\alpha\pi}{2}\right) |\lambda|^{-\alpha-1}. \quad (13.3.13)$$

The distribution has a smooth quadratic maximum at $\lambda = 0$ while for large λ it has heavy power-law tails with the same power as the corresponding stable law for real variables. For $\alpha = 2$ the tails disappear.

The procedure described above to derive the free probability density $\rho(\lambda)$ of free stable laws is solely based on relations between the R -transform, the resolvent and the density. It does not involve any matrix calculations. Free random matrices appear as a representation of these free random variables and correspond to infinitely large random matrices with a given eigenvalue density $\rho(\lambda)$ and with a rotationally invariant probability measure. Actually there are many different matrix realizations of the same free random variable. We shall below discuss two simplest ones which have completely different microscopic properties. The most natural realization is a matrix generated by the uniform angular randomization ODO^T of a large diagonal matrix D of size $N \rightarrow \infty$ which has i.i.d. random variables on the diagonal. If we choose the p.d.f. of the diagonal elements as $\rho(\lambda)$, which corresponds to a free stable law, we obtain a matrix realization of a free random variable which is stable under addition. Clearly, the eigenvalues of this matrix are by construction uncorrelated. We can now independently generate many such matrices $O_i D_i O_i^T$, $i = 1, \dots, K$. Due to the stability also the following sum

$$A_{\boxplus K} = \frac{1}{K^{1/\alpha}} \sum_{i=1}^K O_i D_i O_i^T \quad (13.3.14)$$

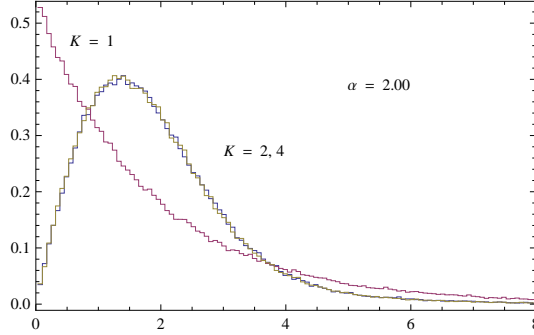


Figure 13.2: Level spacing histograms of the matrices $A_{\boxplus K}$ (13.3.14) of size 200×200 obtained from diagonal matrices D_i whose eigenvalues were generated independently from a Wigner semicircle. For $K = 1$ the histogram properly reflects the Poissonian nature of eigenvalues of a single matrix D_i . For $K > 1$ (already for $K = 2$) the histogram has a shape of the Wigner surmise characteristic for a typical eigenvalue repulsion.

has exactly the same eigenvalue density $\rho(\lambda)$ as each term in the sum, so it is a representation of the same free random variable. However the sum $A_{\boxplus K}$ belongs to a different microscopic class since its eigenvalues repel each other contrary to eigenvalues of each matrix in the sum which are uncorrelated by construction. Already for $K = 2$ one observes a standard repulsion characteristic for invariant ensembles as illustrated in Figure 13.2. It is tempting to conjecture that for $K \rightarrow \infty$ the sum $A_{\boxplus K}$ becomes a random matrix which maximizes entropy for the stable eigenvalue density $\rho(\lambda)$. The probability measure for such matrices is known to be rotationally invariant $d\mu(A) = DA \exp -N \text{tr} V(A)$ and to have a potential $V(\lambda)$ which is related to $\rho(\lambda)$ as [Bal68]

$$V'(\lambda) = 2P.V. \int d\zeta \frac{\rho(\zeta)}{\lambda - \zeta} = G(\lambda + i0^+) + G(\lambda - i0^-) . \quad (13.3.15)$$

For free stable laws the resolvent $G(\lambda)$ is given by a solution of (13.3.11). In particular, for $\alpha = 2$ the last equation gives a quadratic potential $V(\lambda) = \lambda^2/2$, while for $\alpha = 1$ and $\beta = 0$ a logarithmic one $V(\lambda) = \ln(\lambda^2 + 1)$. The potential can also be determined for other values of α . Generally, for large $|\lambda| \rightarrow \infty$ it behaves as

$$V(\lambda) = 2 \ln \lambda - 2\alpha^{-1} \text{Re} (b\lambda^{-\alpha}) + \dots \quad (13.3.16)$$

For maximal entropy random matrices one can also find the joint probability which takes a standard form

$$\rho(\lambda_1, \dots, \lambda_N) = C e^{-N \sum_i V(\lambda_i)} \prod_{i < j} (\lambda_i - \lambda_j)^\beta \quad (13.3.17)$$

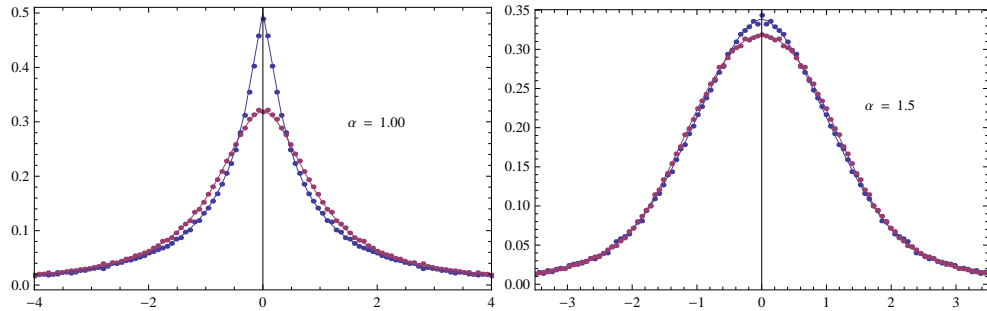


Figure 13.3: Eigenvalue density for Wigner-Lévy matrices and corresponding free Lévy matrices for $\alpha = 1.00$ (left) and $\alpha = 1.50$ (right) are represented by solid line. The corresponding Monte-Carlo histogram for a free sum (13.3.18) of $K = 32$ Wigner-Lévy matrices of size $N = 200$ is shown. For comparison we chose the range of the Wigner-Lévy matrix to be $R = (\Gamma(1 + \alpha))^{-1/\alpha}$ since then the asymptotic behavior of the eigenvalue density for Wigner-Lévy matrices (13.2.10) is identical as for the standardized free random variables (13.3.13).

where C is a normalization and β as usual is equal 1 or 2 for orthogonal or unitary invariant ensemble respectively. For free stable laws the potential $V(\lambda)$ assumes however a highly non-standard non-polynomial form [Bur02].

The stable laws in a free probability play a similar role as the corresponding stable laws in classical probability where a sum of i.i.d. centered random variables $s_n = (x_1 + \dots + x_n)/n^{1/\alpha}$ is known to become an α -stable random variable for $n \rightarrow \infty$. We expect a similar effect for random matrices [Hia00]. For example, if one generates a sequence of K independent Wigner-Lévy matrices (A_1, A_2, \dots, A_K) and a sequence independent random orthogonal matrices (O_1, O_2, \dots, O_K) and one forms a sum

$$A_{\boxplus K} = \frac{1}{K^{1/\alpha}} \sum_{i=1}^K O_i A_i O_i^T \quad (13.3.18)$$

in analogy to (13.3.14) one expects that for large K the sum will become a free random matrix with an eigenvalue density given by a free α -stable law [Bur07]. Actually in practice the convergence to the limiting distribution is very fast. We observe that already for K of order 10 and N of order 100 the eigenvalue density of the matrix $A_{\boxplus K}$ does not significantly differ from the stable density $\rho(\lambda)$ for free random variables (see Figure 13.3). If we skipped random rotations in (13.3.18) and just added many Wigner-Lévy matrices $(A_1 + \dots + A_K)/K^{1/\alpha}$ we would obtain a Wigner-Lévy matrix. As we mentioned already, the independence of the matrices A_i itself is not sufficient to make the addition free and only the rotational randomization brings the sum to the universality

class of free random variables. The comparison of the eigenvalue density of the Wigner-Lévy and the corresponding free Lévy matrices is shown in Figure 13.3.

13.4 Heavy tailed deformations

In this section we discuss ensembles of random matrices obtained from the standard ensembles by a reweighting of the probability measure. We will begin by sketching the idea for real random variables where the procedure is simple and well known and then we will generalize it to random matrices. We will in particular concentrate on heavy-tailed deformations of the probability measure of the Wishart ensemble which are applicable to the statistical multivariate analysis of heavy tailed data.

Consider a random variable x constructed as a product $x = \sigma\xi$ where ξ is a normally distributed variable $\mathcal{N}(0, 1)$ and $\sigma \in (0, \infty)$ is an independent random variable representing a fluctuating scale factor having a p.d.f. $f(\sigma)$. One can think of x as a Gaussian variable $\mathcal{N}(0, \sigma^2)$ for which the variance itself is a random variable. Obviously the p.d.f. of x can be calculated as

$$p(x) = \int_0^\infty d\sigma f(\sigma) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \quad (13.4.1)$$

Choosing appropriately the frequency function $f(\sigma)$ one can thus model the p.d.f. of x . For example for [Tos04, Ber04, Abu05]

$$f(\sigma) = \frac{1}{\sigma} \frac{2}{\Gamma(\frac{\alpha}{2})} \left(\frac{a^2}{2\sigma^2} \right)^{\frac{\alpha}{2}} e^{-\frac{a^2}{2\sigma^2}} \quad (13.4.2)$$

where a is a constant, the integral (13.4.1) takes the following form

$$p(x) = \frac{1}{a\sqrt{\pi}} \frac{1}{\Gamma(\frac{\alpha}{2})} \int_0^\infty d\zeta \zeta^{\frac{\alpha-1}{2}} e^{-\zeta} e^{-\zeta(\frac{x}{a})^2} = \frac{1}{a\sqrt{\pi}} \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{\alpha}{2})} \left(1 + \left(\frac{x}{a} \right)^2 \right)^{-\frac{\alpha+1}{2}}. \quad (13.4.3)$$

In doing the integral we changed the integration variable⁶ to $\zeta = a^2/(2\sigma^2)$. For large $|x|$ the p.d.f. has a power-law tail $p(x) \sim |x|^{-\alpha-1}$. The variance of the distribution exists only for $\alpha > 2$ and is equal to $a^2/(\alpha - 2)$.

If one applies this procedure to each matrix element independently $A_{ij} = \sigma_{ij}\xi_{ij}$, $i \leq j$, one obtains a Wigner matrix discussed in the first section of this chapter. One can however construct a slightly different matrix $A_N(\sigma)$ whose elements have the same common random scale factor $A_{ij} = \sigma\xi_{ij}$ and differ only by ξ_{ij} which are i.i.d. Gaussian random variables $\mathcal{N}(0, 1)$. In the limit $N \rightarrow \infty$ the eigenvalue density of the matrix $A_N(\sigma)/\sqrt{N}$ converges to the Wigner semicircle law $\rho_\sigma(\lambda) = \sqrt{4\sigma^2 - \lambda^2}/(2\pi\sigma^2)$. The scale factor σ changes

⁶The variable ζ has a χ^2 distribution.

however from matrix to matrix with the frequency $f(\sigma)$ so in analogy to (13.4.1) the effective eigenvalue density in the ensemble of matrices is given by the average of the semicircle law over σ [Boh08]

$$\rho(\lambda) = \int d\sigma f(\sigma) \rho_\sigma(\lambda) = \int_{\frac{\lambda}{2}}^{\infty} d\sigma f(\sigma) \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - \lambda^2}. \quad (13.4.4)$$

In particular, for the frequency function (13.4.2) we obtain the following eigenvalue density [Ber04]

$$\rho(\lambda) = \frac{1}{a} \frac{\sqrt{2}}{\pi \Gamma(\frac{a}{2})} \int_0^{\frac{2a^2}{\lambda^2}} d\zeta \zeta^{\frac{a-1}{2}} e^{-\zeta} \sqrt{1 - \frac{\zeta \lambda^2}{2a^2}} \quad (13.4.5)$$

which has power-law tails, $\rho(\lambda) \sim |\lambda|^{-\alpha-1}$, with the same power as the p.d.f. (13.4.3) for matrix elements. One can generalize the result to other frequency functions [Mut05].

Actually exactly the same strategy can be applied to calculate the joint probability since fluctuations of matrix elements are modified by a common scale factor $\rho(\lambda_1, \dots, \lambda_N) = \int d\sigma f(\sigma) \rho_\sigma(\lambda_1, \dots, \lambda_N)$, where ρ_σ is given by (13.3.17) with $V(\lambda) = \lambda^2/(2\sigma^2)$. From the joint probability one can then derive microscopic properties of matrices, including the microscopic correlation functions and the distribution of the largest eigenvalue [Boh09].

One can use this procedure to deform probability measures of other matrix ensembles as well. In what follows we concentrate on deformations of the Wishart ensemble. The probability measure for a standardized Wishart ensemble of real matrices⁷ is given by

$$d\mu_*(\xi) = (2\pi)^{-\frac{NT}{2}} e^{-\frac{1}{2}\text{tr} \xi \xi^T} D\xi \quad (13.4.6)$$

where ξ is a rectangular matrix ξ_{it} , $i = 1, \dots, N$, $t = 1, \dots, T$ and $D\xi = \prod_{i,t}^{N,T} d\xi_{it}$. The eigenvalue density of the matrix $(1/T)\xi\xi^T$ is known to converge to the Marčenko-Pastur law $\rho_*(\lambda) = \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}/(2\pi r\lambda)$, where $\lambda_{\pm} = (1 \pm r)^2$ and $r = N/T$ in the limit $N \rightarrow \infty$, $r = \text{const}$ [Mar67]. The elements of the matrix ξ represent normally distributed fluctuations of uncorrelated random numbers with unit variance. Using now the reweighting method we can consider a matrix $A_{it} = \sigma \xi_{it}$ with a common fluctuating scale factor being an independent random variable with a p.d.f. $f(\sigma)$. The effective probability measure can be easily derived from (13.4.6) and reads

$$d\mu(A) = DA \int d\sigma f(\sigma) \sigma^{-NT} e^{-\frac{1}{2\sigma^2}\text{tr} AA^T} \quad (13.4.7)$$

⁷For complex matrices the corresponding measure reads $d\mu_*(\xi) = \pi^{-NT} e^{-\text{tr} \xi \xi^\dagger} D\xi$ where $D\xi = \prod_{it}^{NT} d\text{Re} \xi_{it} d\text{Im} \xi_{it}$

where the factor σ^{-NT} comes from the change of variables in the measure $DA = \sigma^{NT} D\xi$. In particular for the frequency function (13.4.2) this gives the measure of a multivariate Student's distribution⁸

$$d\mu(A) = DA \frac{\Gamma(\frac{\alpha+NT}{2})}{(a\sqrt{\pi})^{NT} \Gamma(\frac{\alpha}{2})} \left(1 + \frac{\text{tr} AA^T}{a^2}\right)^{-\frac{\alpha+NT}{2}}. \quad (13.4.8)$$

The eigenvalue density of the matrix $(1/T)XX^T$, where X is generated with the probability measure given above can be obtained by the same reweighting method as before. The eigenvalue density of the Gaussian part in (13.4.7) is $\rho_\sigma(\lambda) = \rho_*(\lambda/\sigma^2)/\sigma^2 = \sqrt{(\sigma^2\lambda_+ - \lambda)(\lambda - \sigma^2\lambda_-)}/(2\pi r\sigma^2\lambda)$. It has to be reweighted with the frequency function (13.4.2) $\rho(\lambda) = \int d\sigma f(\sigma)\rho_\sigma(\lambda)$. The result reads [Bur06]

$$\rho(\lambda) = \frac{(\frac{\alpha}{2})^{\alpha/2}}{2\pi r \Gamma(\frac{\alpha}{2})} \lambda^{-\alpha/2-1} \int_{\lambda_-}^{\lambda_+} \sqrt{(\lambda_+ - \zeta)(\zeta - \lambda_-)} e^{-\frac{\alpha\zeta}{2\lambda}} \zeta^{\alpha/2-1} d\zeta. \quad (13.4.9)$$

The support of this eigenvalue distribution is infinite. The exponent $\alpha/2$ of the tail is a half of the exponent α of p.d.f. for the matrix elements X as one can expect for a matrix $(1/T)XX^T$ which is a “square” of X . The reweighting can be applied to derive the corresponding joint probability function and to determine the microscopic correlations of the deformed Wishart ensembles [Ake08].

This result can be generalized in a couple of ways. One can change the frequency function $f(\sigma)$ [Abu09] but one can also change the relation between the A and ξ matrix from $A_{it} = \sigma\xi_{it}$ to, for instance, $A_{it} = \sigma \sum_j S_i O_{ij} \xi_{jt}$ where O_{ij} is a rotation matrix and S_i is a vector of positive numbers. The matrix O and the vector S are fixed in this construction and the only fluctuating elements are ξ_{it} which are i.i.d. $\mathcal{N}(0, 1)$ and σ which is an independent random variable with a given p.d.f. $f(\sigma)$, as before. The interpretation of the construction is clear. The factors S_i 's change the scale of fluctuations and the matrix O rotates the main axes. If one applies it to (13.4.6) one will obtain a deformed measure (13.4.8) where $\text{tr} AA^T$ will be substituted by $\text{tr} AC^{-1}A^T$. The matrix $C_{ij} = \sum_k O_{ik} S_k^2 O_{kj}$ introduces explicit correlations between the degrees of freedom. In a similar way one can introduce correlations $C_{tt'}$ between A_{it} and $A_{it'}$ at different times t and t' [Bur06].

Another interesting generalization of the reweighting procedure is to consider a matrix $A_{it} = \sigma_i \xi_{it}$ where now the scale factors σ_i are independent random variables for each row [Bir07b]. This case corresponds to a Wishart ensemble with a fluctuating covariance matrix $C_{ij} = \delta_{ij} \sigma_i^2$ where σ_i are i.i.d.

⁸An analogous expression [Tos04, Ber04] can be obtained for GOE and GUE Wigner matrices reweighted with the frequency function (13.4.2).

random variables with a given p.d.f. $f(\sigma)$. This problem can be solved analytically thanks to an explicit relation between the Greens functions and eigenvalue densities of the matrices C and $(1/T)XX^T$ [Bur04].

The idea of reweighting is quite general. So far we have described reweighting through the scale parameter σ but one can use other quantities as a basis for the reweighting scheme as well. For example one can use the trace $t = \text{tr}XX^T$ of the whole matrix. In this scheme the idea is to calculate quantities for the ensemble with a probability measure⁹ $d\mu_t(X) = DX\delta(t - \text{tr}XX^T)$ and then to reweight them using a frequency function $g(t)$ to obtain the corresponding values for the ensemble with the measure $d\mu(X) = DXg(\text{tr}XX^T)$ [Ake99, Abu05]. It turns out that the first step, that is the calculations for the fixed trace ensemble, can be done analytically so also this procedure gives a practical recipe to handle “non-standard” ensembles. Of course it works only for ensembles for which the measure depends on $\text{tr}XX^T$ as for instance (13.4.8). In particular it can be applied to the multivariate Wishart-Student ensembles [Bur06].

13.5 Summary

Heavy-tailed random matrices is a relatively new branch of random matrix theory. In this chapter we discussed several matrix models belonging to this class and presented methods of integrating them. We believe that the models, methods and concepts can be applied to many statistical problems where non-Gaussian effects play an important role.

ACKNOWLEDGEMENTS

We thank G. Akemann, J.P. Bouchaud, A. Görlich, R.A. Janik, A. Jarosz, M.A. Nowak, G. Papp, P. Vivo, B. Waclaw and I. Zahed for many interesting discussions. This work was supported by the Marie Curie ToK project “COCOS”, No. MTKD-CT-2004-517186, the EC-RTN Network “ENRAGE”, No. MRTN-CT-2004-005616 and the Polish Ministry of Science Grant No. N N202 229137 (2009-2012).

References

- [Abu05] A. Y. Abul-Magd, Phys. Rev. E **71** (2005) 066207.
- [Abu09] A.Y. Abul-Magd, G. Akemann, P. Vivo, J. Phys. A: Math. Theor. **42** (2009) 175207.

⁹Alternatively one can use $d\mu'_t(X) = \theta(\text{tr}XX^T - t)DX$, where $\theta(x)$ is the Heaviside step function [Ber04].

- [Ake99] G. Akemann, G.M. Cicuta, L. Molinari, G. Vernizzi, Phys. Rev. E **59** (1999) 1489; Phys. Rev. E **60** (1999) 5287.
- [Ake08] G. Akemann and P. Vivo, J. Stat. Mech. (2008) P09002.
- [Bal68] R. Balian, Nuovo Cimento B **57** (1968) 183.
- [Ber93] H. Bercovici and D. Voiculescu, Ind. Univ. Math. J. **42** (1993) 733.
- [Ber99] H. Bercovici and V. Pata, Annals of Mathematics **149** (1999) 1023; Appendix by P. Biane.
- [Ber04] A.C. Bertuola, O. Bohigas, and M.P. Pato, Phys. Rev E **70** (2004) 065102(R).
- [Bir07a] G. Biroli, J.-P. Bouchaud, and M. Potters, Europhys. Lett. **78** (2007) 10001.
- [Bir07b] G. Biroli, J.-P. Bouchaud, and M. Potters, Acta Phys. Pol. B **38** (2007) 4009.
- [Boh08] O. Bohigas, J.X. de Carvalho and M.P. Pato, Phys. Rev. E **77** (2008) 011122.
- [Boh09] O. Bohigas, J.X. de Carvalho and M.P. Pato, Phys. Rev. E **79** (2009) 031117.
- [Bou97] J.-P. Bouchaud and M. Mézard, J. Phys. A, Math. Gen. **30** (1997) 7997.
- [Bur02] Z. Burda, R.A. Janik, J. Jurkiewicz, M.A. Nowak, G. Papp and I. Zahed, Phys. Rev. E **65** (2002) 021106.
- [Bur04] Z. Burda, A. Görlich, A. Jarosz, J. Jurkiewicz, Physica A **343** (2004) 295.
- [Bur06] Z. Burda, A. Görlich, B. Waclaw, Phys. Rev. E **74** (2006) 041129.
- [Bur07] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp and I. Zahed, Phys. Rev. E **75** (2007) 051126; arXiv:cond-mat/0602087.
- [Ciz94] P. Cizeau and J.P. Bouchaud, Phys. Rev. E **50** (1994) 1810.
- [Fel71] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, 3rd Edition, New York 1971.
- [Gne68] B.V. Gnedenko, A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Revised Edition Addison-Wesley, Cambridge 1968.

- [Hia00] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, Am. Math. Soc., Providence 1992.
- [Mar67] V.A. Marčenko and L. A. Pastur, Math. USSR-Sb, **1**, (1967) 457.
- [Meh04] M.L. Mehta, *Random Matrices*, Academic Press, 3rd Edition, London 2004.
- [Mut05] K.A. Muttalib and J.R. Klauder, Phys. Rev. E **71**, (2005) 055101(R).
- [Nic06] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, London Mathematical Society Lecture Note Series, vol. 335, Cambridge University Press, 2006.
- [Nol10] J.P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*, Birkhäuser, Boston 2010; <http://academic2.american.edu/~jpnolan>.
- [Pas72] L.A. Pastur: Teor. Mat. Fiz., **10** (1972) 102.
- [Rao06] N.R. Rao, *RMTool: A random matrix and free probability calculator in MAT-LAB*; <http://www.mit.edu/~raj/rmtool/>.
- [Sos04] A. Soshnikov, Elect. Comm. in Probab. **9** (2004) 82.
- [Spe94] R. Speicher, Math. Ann. **298** (1994) 611.
- [Tos04] F. Toscano, R.O. Vallejos and C. Tsallis, Phys. Rev. E **69** (2004) 066131.
- [Tra94] C.A. Tracy and H. Widom, Commun. Math. Phys. **159** (1994) 151.
- [Tra96] C.A. Tracy and H. Widom, Commun. Math. Phys. **177** (1996) 724.
- [Voi85] D.V. Voiculescu, in *Operator algebras and their connections with topology and ergodic theory*, (Busteni, 1983), Lecture Notes in Math. Series, vol. **1132**, 556, Springer, New York 1985.
- [Voi91] D.V. Voiculescu, Invent. Math. **104** (1991), 201220.
- [Voi92] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, Am. Math. Soc., Providence 1992.